SOME UNIFYING THOUGHTS ON THE ANALYSIS OF VARIANCE*

S. R. Searle

Biometrics Unit, Cornell University, Ithaca, New York

BU-562-M June, 1975

Abstract

Analysis methods for the general linear model $\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{e}$ cover many more situations than they are often given credit for: e.g., designed experiments, survey data, regression, polynomial fitting and covariance. This umbrella-like aspect of the general linear model is important in 2 ways. First, in teaching, where it provides a unified procedure adaptable to many topics that are all too frequently taught as isolated entities; and second, in analyzing real data, where it is always a back-up procedure for developing analysis methods for models that a user may not have previously encountered or for which "recipes" are not readily available.

---

* Invited Paper presented at the American Statistical Association Meetings, Atlanta, Georgia, August 1975.

SOME UNIFYING THOUGHTS ON THE ANALYSIS OF VARIANCE

S. R. Searle
Biometrics Unit, Cornell University, Ithaca, New York

BU-562-M June, 1975

Introduction

Analysis of variance is among the most widely used of statistical methods, and is employed in many different situations. Yet in the teaching of statistics there is undoubtedly much duplication and overlapping of the topics involved. To the extent that duplication means repetition and hence reinforcement of ideas for the student, then certainly some degree of duplication is to be applauded. That is good teaching. But it is not good teaching when duplication and the overlapping of topics is not explained but is left to be gleaned by anology and implication. This certainly occurs, and leads students towards an attitude of thinking that each topic encountered is isolated from others. As a result they may end up feeling, for example, that the analyses of variance for split plot experiments, for regression and for covariance are unrelated activities. Not only is such an attitude wrong but it makes the learning and understanding of these analyses much more difficult than need be. In contrast, what should be emphasized is the unifying aspects of general linear model (GLM) theory. Under its umbrella come all the individual analyses of variance of experimental design data, of survey data, of regression, of polynomial fitting and of covariance. Understanding a few basic results in GLM theory enables one to view, and carry out, all these analyses as special cases of one unified procedure, without having to think of each one as something isolated and separate from the others.

Fisher and designed experiments

Consider a randomized complete block (rcb) experiment with n observations on each of t treatments in each of b blocks. Let

$$y_{ijk} \equiv \text{k'th observation in i'th treatment in j'th block;} \tag{1}$$

we use the familiar dot and bar notation for totals and means, e.g.,

$$y_{i..} = \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk} \quad \text{and} \quad \bar{y}_{i..} = y_{i..}/bn \ . \tag{2}$$

As Urquhart *et al.* [1972] have so eloquently pointed out, the analysis of variance of data such as these was first described by Fisher without any reference to today's familiar linear model for a rcb. The essentials of what Fisher did consisted first of observing the algebraic identity

$$\sum_{i=1}^{t} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(y_{ijk} - \bar{y}_{...}\right)^2$$

$$\equiv bn \sum_{i=1}^{t} \left(\bar{y}_{i..} - \bar{y}_{...}\right)^2 + tn \sum_{j=1}^{b} \left(\bar{y}_{.j.} - \bar{y}_{...}\right)^2 + n \sum_{i=1}^{t} \sum_{j=1}^{b} \left(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}\right)^2$$

$$+ \sum_{i=1}^{t} \sum_{j=1}^{b} \sum_{k=1}^{n} \left(y_{ijk} - \bar{y}_{ij.}\right)^2 \ . \tag{3}$$

He then showed, under normality, i.e., with the y_{ijk}'s being independently and normally distributed with uniform variance σ^2, that the four sums of squares on the right-hand side of (3) are independently distributed as multiples of χ^2-distributions, and from this the F-statistics were derived and used as we have them today. Furthermore, the familiar analysis of variance (ANOVA) table was developed, primarily to provide a convenient summary of all the calculations required.

Table 1: ANOVA Table for RCB Design

Source of Variation	d.f.	Sum of Squares	Mean Square	F-statistic
Treatments	t-1	$SST=bn\sum_{i=1}^{t}\left(\bar{y}_{i..}-\bar{y}_{...}\right)^2$	$MST=SST/(t-1)$	$FF_T=MST/MSE$
Blocks	b-1	$SSB=tn\sum_{j=1}^{b}\left(\bar{y}_{.j.}-\bar{y}_{...}\right)^2$	$MSB=SSB/(b-1)$	$F_B=MSB/MSE$
Interaction	(t-1)(b-1)	$SSTB=n\sum_{i=1}^{t}\sum_{j=1}^{b}\left(\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...}\right)^2$	$MSTB=SSTB/N'$ with $N'=(t-1)(b-1)$	$F_{TB}=MSTB/MSE$
Residual	tb(n-1)	$SSE=\sum_{i=1}^{t}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk}-\bar{y}_{ij.}\right)^2$	$MSE=SSE/tb(n-1)$	
Total	tbn-1	$SSO=\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk}-\bar{y}_{...}\right)^2$		

Note that nowhere in the preceding discussion is any use made of a model equation for y_{ijk} such as we have it today, e.g.

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} . \qquad (4)$$

As Urquhart et al. [1973] so nicely describe, the use of this sort of thing post-dated Fisher's original work by some years. For those who to-day hang their hat on model equations like (4), it may be revealing that in order to have the ANOVA that is so familiar to us all, no model equation is necessary. The total variation among the observations is "clearly" represented by $SSO = \sum_{i=1}^{t}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(y_{ijk} - \bar{y}_{...}\right)^2$, the lefthand side of (3). Furthermore, the righthand side of (3) represents a partitioning of SSO into four terms each of which has a similar kind of meaning, e.g., $SST = bn\sum_{i=1}^{t}\left(\bar{y}_{i..} - \bar{y}_{...}\right)^2$ represents

variation among treatment means. Thus, not only is (3) an algebraic identity but it seems to be a very reasonable partitioning of the total sum of squares. Furthermore, through assuming normality we have the independence and χ^2 properties of the sums of squares in this partitioning and hence evolve the F-statistics —with the calculations all conveniently summarized in the ANOVA table.

There is no word of a model equation like (4) in the early development of the ANOVA table. The formal model came much later in time, as a crutch designed for helping to understand analysis of variance. Unfortunately we are now coming to see that that crutch has sometimes failed us, and that maybe we need to reassess its utility.

Complex designed experiments

In the years since Fisher's development of ANOVA there has been enormous expansion in the variety of designed experiments. Stemming from Latin squares and their Graeco relatives, from lattice designs, incomplete blocks and fractional replicates we now have a vast array of experimental designs, some of them quite complicated. For each design there is an appropriate ANOVA table whose main entries are sums of squares written as summations like those of Table 1.

Analyzing the data from any of these experiments requires 2 major steps. The first is to understand how the data have been gathered, i.e., how the experiment was carried out. Knowing this, the second step is to know, or be able to develop, an ANOVA for the experiment. Those who are experts in the subject of experimental design usually excel in both of these steps. But many of us are not experts; as one such I never cease to marvel at the speed with which an experimental design expert, given the description of an experiment,

can toss off the ANOVA table for a complex design and, what's more, get it correct!

As substitute for the experience that the expert has, there are the well known rules for developing ANOVA tables for complete factorial experiments such as those developed by Lum [1954] and Schultz [1955] (see also Searle [1971, Sec. 9.6]). But these rules do not encompass the myriad of designs that are in any sense fractional replicates; nor do they encompass unequal-subclass-numbers data, or unbalanced, messy or survey data as they have also been called. For these we must either know the ANOVA methods explicitly (which the design expert does for complex designs), or we must have some general procedure that can provide us with the explicit methods. We do: the general linear model procedure.

Modeling with a linear model

As has been said, the first step in analyzing data is to understand how the data have been gathered and from this to postulate a model. This process, which is usually no easy task, involves both the statistician and the investigator whose data are to be analyzed. Although it is the crux of the whole art of modeling it is also the aspect of modeling to which little attention is given in teaching or in textbooks. This is probably so because the art of modeling is somewhat nebulous, it is an art, and to that extent is not easy to put in words. As a result, most writings begin at the point "let us assume such and such a model" and they proceed to follow the consequences. We shall be no exception here: our objective is to highlight the umbrella-like nature and unifying features of the general linear model vis à vis several special cases of it. Nevertheless, one must never overlook the implicit difficulties

that hide behind our starting point of "let us assume the following model". This phrase, in the presence of a body of data, demands a lot of hard thinking.

The factors that play a part in designing an experiment will usually be those whose effects are to be studied in analyzing the data from the experiment. With survey data, decisions have to be made as to which of the recorded factors shall be studied and how. In either case, denoting a datum (or a transformed datum) by $y_{ij\cdots pr}$, when dealing with linear models we postulate the form

$$y_{ij\cdots pr} = E\left(y_{ij\cdots pr}\right) + \epsilon_{ij\cdots pr}$$

where $y_{ij\cdots pr}$ is the r'th observation in the i'th, j'th, $\cdots$, p'th levels of the first, second, $\cdots$, factors respectively. In this context $y_{ij\cdots pr}$ is considered as a random sample from the population defined by the subscripts i, j, $\cdots$, p and $E\left(y_{ij\cdots pr}\right)$ is its expected value over that population. This we denote by $\mu_{ij\cdots p}$, so getting

$$y_{ij\cdots pr} = \mu_{ij\cdots p} + \epsilon_{ij\cdots pr} .$$

This is the model equation; it is not the model. Specification of the model requires description of the properties of $\mu_{ij\cdots p}$ and $\epsilon_{ij\cdots pr}$.

To make discussion easier we use only 3 subscripts and consider simply

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk} . \qquad (5)$$

This is the model equation for just a 2-factor model: y_{ijk} is the k'th observation in the cell defined by the i'th level of one factor and the j'th level of the other factor; μ_{ij} is the mean of the population corresponding to that cell and ϵ_{ijk} is the deviation of y_{ijk} from μ_{ij}, a deviation that is assumed to be random. This is just the general kind of model discussed by Urquhart et al. [1972].

The problem with (5) is that we generally like to be more specific about the cell mean than denoting it simply by μ_{ij}. This desire for specificity leads, for example in the rcb case, to defining μ_{ij} as

$$\mu_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} \tag{6}$$

and so getting the model equation we had in (4)

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} . \tag{7}$$

A complete description of the model requires specifying which effects in (7) are random and which are fixed, and for the random effects we must define certain properties of their distributions, e.g., mean and variance-covariance structure.

Models postulated in the manner of (7) are called linear models because they are linear functions of the parameters involved, the τ_i's, β_j's and $(\tau\beta)_{ij}$'s. They are also appealing because they coincide rather well with intuition: there is a mean μ and an effect for each factor and effects for interactions between factors. Models can also include covariates as, for example,

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + b_1 x_{1,ijk} + b_2 x_{2,ijk} + \epsilon_{ijk} . \tag{8}$$

If the only terms in a model are covariates (with or without a mean μ) then we have a regression model. In all cases, an obvious advantage of model equations like (7) and (8) is that writing them down helps one think seriously about which factors (and covariates) are to be studied, and how. The equations play a similar role in using computer programs. Software packages for ANOVA need input specifying the factors being used, and often require the specification to be akin to the notation used in (7) and (8), including indication, of course, as to which effects are fixed and which are random.

When ϵ_{ijk} is the only random term in equations such as (7) and (8), models based thereon have one well-known problem. They are over-parameterized. Thus in (7) with at least one observation in each of the tb cells there are $1 + t + b + tb$ parameters in the model (over and above any variance-covariance parameters), but there are only tb cell means from which to attempt estimating them as linear combinations of those means. Obviously this is an impossible task. And when some of the cells have no data in them then it is even more of an impossible task.

One way of overcoming the over-parameterization problem, in the case of all cells filled, is to define the parameters in (7) subject to $1 + t + b + tb - tb = 1 + t + b$ restrictions on them. A popular set of such restrictions is

$$\sum_{i=1}^{t} \tau_i = 0, \quad \sum_{j=1}^{b} \beta_j = 0, \quad \sum_{i} (\tau\beta)_{ij} = 0 \text{ for all } j, \text{ and } \sum_{j} (\tau\beta)_{ij} = 0 \text{ for all } i .$$

Although these kinds of restrictions are eminently useful and reasonable when all cells have the same number of observations, and in designed experiments generally, the same is not quite so true for survey data where scattered cells exist having no data at all. Furthermore, these restrictions are often confused with constraints on solutions to normal equations, constraints imposed solely for the purpose of deriving solutions. The whole topic of these restrictions and constraints can be fraught with confusion and we need say no more here other than that they can be handled within the framework of general linear model (GLM) theory.

Unifying features of the GLM

Stripped of many surrounding details the GLM can be summarized as follows. Suppose data to be analyzed are N observations y_i for i = 1, ···, N. The equation of the general linear model for y_i can be represented using

$$E(y_i) = \sum_{j=0}^{k} x_{ij} b_j \qquad (9)$$

i.e., as

$$y_i = \sum_{j=1}^{k} x_{ij} b_j + \epsilon_i \ . \qquad (10)$$

(9) represents the mean value $E(y_i)$ as the linear combination $\sum_{j=0}^{k} x_{ij} b_j$ of k + 1 unknown parameters b_j for j = 0, 1, ···, k, and (10) is the model equation. The coefficients x_{ij} of these parameters in (9) can take several forms:

(i) For j = 0, say, each x_{i0} for i = 1, ···, N may be unity; thus the x_{i0}'s are all the dummy variable 1, a situation that corresponds to $b_0 \equiv \mu$ in models like (7) and (8).

(ii) For some values j the x_{ij}'s may be dummy variables, 0 or 1, corresponding to the associated b_j's being effects for factors and interactions.

(iii) For some values j the x_{ij}'s may be variables other than the y_i's. These variables can be observed, or pre-assigned, values (e.g. observed height, or pre-assigned time intervals); or they may be linear or non-linear functions of such observations.

The 7 combinations in which these 3 kinds of x-variables can occur in a model are shown in Table 2. (See next page.) Combination A is the simplest and least useful model, although it <u>is</u> the model for which the correction for the mean, $N\bar{y}_.^2$, is the reduction in sum of squares. B, the factors-and-interaction model, is exemplified by (4). C is linear regression, and it is worth emphasizing that linear here continues to mean linearity of the parameters b_j in the

Table 2: The 7 Combinations of the 3 Kinds of x-Variable in a Linear Model

Combination	x-variables			Model
	mean (i)	factors (ii)	covariates (iii)	
A	✓			Mean only: $y_i = \mu + \epsilon_i$
B	✓	✓		Factors and interactions
C	✓		✓	Regression
D			✓	Regression through origin
E	✓	✓	✓	Covariance
F		✓		B, without mean μ
G		✓	✓	E, without mean μ

model (9). It does not mean linearity of the observed x-variables; indeed non-linearity of these is specifically provided for in (iii). In particular, if $x_{ij} = t^j$ for all i, for j = 0, 1, 2, ···, k then the model is that of fitting a polynomial in t. Cases B, C and E are the three most important combinations in Table 2, of which E, covariance, is really no more than a combination of B and C. This is evident from the table, which shows that covariance has x-variables (i), (ii) and (iii) whereas factors-and-interactions models have (i) and (ii) and regression has (i) and (iii). D is a special case of C, namely regression through the origin and F and G are slight reparameterizations of B and E that exclude a general mean μ.

For any of the combinations in Table 2 let $\underset{\sim}{y}$ be the vector of the y_i's in (10), $\underset{\sim}{\beta}$ the vector of b_j's, $\underset{\sim}{X}$ the N × (k + 1) matrix of x_{ij}'s and $\underset{\sim}{e}$ the vector of ϵ_i's. Then the model (10) has the well known general form

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{e} \; . \qquad (11)$$

In this notation, properties of the GLM and its general procedures can be summarized as follows.

Estimation Normal equations arising from the method of least squares for estimating $\underset{\sim}{\beta}$ are

$$\underset{\sim}{X}'\underset{\sim}{X}\underset{\sim}{\beta}^{o} = \underset{\sim}{X}'\underset{\sim}{y} \tag{12}$$

with solution

$$\underset{\sim}{\beta}^{o} = (\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{y} \tag{13}$$

where $(\underset{\sim}{X}'\underset{\sim}{X})^{-}$ is any generalized inverse of $\underset{\sim}{X}'\underset{\sim}{X}$ satisfying

$$\underset{\sim}{X}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{X} = \underset{\sim}{X}'\underset{\sim}{X} \ . \tag{14}$$

Although there are many matrices $(\underset{\sim}{X}'\underset{\sim}{X})^{-}$ satisfying (14), it is true for each of them that

$$\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}' \text{ is invariant to } (\underset{\sim}{X}'\underset{\sim}{X})^{-} \ . \tag{15}$$

As a result, although $\underset{\sim}{\beta}^{o}$ is not invariant to the choice of $(\underset{\sim}{X}'\underset{\sim}{X})^{-}$, $\hat{\underset{\sim}{y}} = \underset{\sim}{X}'\underset{\sim}{\beta}^{o}$ is invariant. So also is the vector of residuals $\underset{\sim}{y} - \hat{\underset{\sim}{y}}$.

Variance estimation In the model equation (11) the mean and variance of $\underset{\sim}{e}$ are usually taken to be $\underset{\sim}{0}$ and $\sigma^2\underset{\sim}{I}_N$ respectively. An unbiased estimator of σ^2 based on (13) is then

$$\hat{\sigma}^2 = \frac{(\underset{\sim}{y} - \hat{\underset{\sim}{y}})'(\underset{\sim}{y} - \hat{\underset{\sim}{y}})}{N - r(\underset{\sim}{X})} = \frac{\underset{\sim}{y}'[\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}']\underset{\sim}{y}}{N - r(\underset{\sim}{X})} = \frac{\underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{\beta}^{o\prime}\underset{\sim}{X}'\underset{\sim}{y}}{N - r(\underset{\sim}{X})} \tag{16}$$

where $r(\underset{\sim}{X})$ is the rank of $\underset{\sim}{X}$. Because of (15), $\hat{\sigma}^2$ is also invariant to $(\underset{\sim}{X}'\underset{\sim}{X})^{-}$, i.e. to $\underset{\sim}{\beta}^{o}$.

Estimable functions The best linear unbiased estimator (b.l.u.e.) of an estimable function $\underset{\sim}{q}'\underset{\sim}{\beta}$ of the parameters in $\underset{\sim}{\beta}$ is

$$\widehat{\underset{\sim}{q}'\underset{\sim}{\beta}} = \underset{\sim}{q}'\underset{\sim}{\beta}^{\circ} \tag{17}$$

with variance

$$v\left(\underset{\sim}{q}'\underset{\sim}{\beta}\right) = \underset{\sim}{q}'(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{q}\sigma^2 , \tag{18}$$

both these results also being invariant to the choice of $(\underset{\sim}{X}'\underset{\sim}{X})^{-}$.

Test of hypotheses Under normality assumptions, i.e., $\underset{\sim}{e} \sim N\left(\underset{\sim}{0}, \sigma^2\underset{\sim}{I}_N\right)$, the F-statistic for testing the concordance of the data with the model (11) is

$$\frac{\underset{\sim}{y}'\underset{\sim}{X}\left(\underset{\sim}{X}'\underset{\sim}{X}\right)^{-1}\underset{\sim}{X}'\underset{\sim}{y}}{r(\underset{\sim}{X})} \cdot \frac{N - r(\underset{\sim}{X})}{\underset{\sim}{y}'\left[\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\right]\underset{\sim}{y}} \sim F_{r(\underset{\sim}{X}),N-r(\underset{\sim}{X})} . \tag{19}$$

When $\underset{\sim}{K}'\underset{\sim}{\beta}$ is a set of s linearly independent estimable functions the F-statistic for testing the hypothesis $H = \underset{\sim}{K}'\underset{\sim}{\beta} = \underset{\sim}{m}$, for any pre-assigned vector $\underset{\sim}{m}$ is

$$F(H) = (\underset{\sim}{K}'\underset{\sim}{\beta}^{\circ} - \underset{\sim}{m})'[\underset{\sim}{K}'(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{K}]^{-1}(\underset{\sim}{K}'\underset{\sim}{\beta}^{\circ} - \underset{\sim}{m})/s\hat{\sigma}^2 \sim F_{q,N-r(\underset{\sim}{X})} . \tag{20}$$

Estimability properties of $\underset{\sim}{K}'\underset{\sim}{\beta}$ ensures the non-singularity of $\underset{\sim}{K}'(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{K}$.

A partitioned form When (11) can be rewritten as

$$\underset{\sim}{y} = \underset{\sim}{X}_1\underset{\sim}{\beta}_1 + \underset{\sim}{X}_2\underset{\sim}{\beta}_2 + \underset{\sim}{e} \tag{21}$$

the normal equations (12) become

$$\begin{bmatrix} \underset{\sim}{X}_1'\underset{\sim}{X}_1 & \underset{\sim}{X}_1'\underset{\sim}{X}_2 \\ \underset{\sim}{X}_2'\underset{\sim}{X}_1 & \underset{\sim}{X}_2'\underset{\sim}{X}_2 \end{bmatrix} \begin{bmatrix} \underset{\sim}{\beta}_1^{\circ} \\ \underset{\sim}{\beta}_2^{\circ} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{X}_1'\underset{\sim}{y} \\ \underset{\sim}{X}_2'\underset{\sim}{y} \end{bmatrix} . \tag{22}$$

The reduction in sum of squares due to fitting (21) rather than

$$\underset{\sim}{y} = \underset{\sim}{X}_2\underset{\sim}{\beta}_2 + \underset{\sim}{\epsilon} \tag{23}$$

is then

$$\begin{aligned} R(\underset{\sim}{\beta}_1|\underset{\sim}{\beta}_2) &= R(\underset{\sim}{\beta}_1,\underset{\sim}{\beta}_2) - R(\underset{\sim}{\beta}_2) \\ &= \underset{\sim}{y}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{y} - \underset{\sim}{y}'\underset{\sim}{X}_2(\underset{\sim}{X}_2'\underset{\sim}{X}_2)^{-}\underset{\sim}{X}_2'\underset{\sim}{y} \\ &= \underset{\sim}{\beta}_1^{o\,\prime}\underset{\sim}{Q}\underset{\sim}{\beta}_1^{o} \end{aligned} \tag{24}$$

for

$$\underset{\sim}{Q} = \underset{\sim}{X}_1'\underset{\sim}{X}_1 - \underset{\sim}{X}_1'\underset{\sim}{X}_2(\underset{\sim}{X}_2'\underset{\sim}{X}_2)^{-}\underset{\sim}{X}_2'\underset{\sim}{X}_1 \,. \tag{25}$$

This will also be the numerator of (20) for testing the hypothesis $\underset{\sim}{\beta}_2 = \underset{\sim}{0}$ when rewritten for some $\underset{\sim}{K}'$ in the testable form $\underset{\sim}{K}'\underset{\sim}{\beta} = \underset{\sim}{0}$. Repeated use of this technique in its full detail is what provides facilities for deriving ANOVA tables for complex experiments and for survey data.

Some Special Cases

Designed experiments This is not the place to give lengthy examples of the use of these GLM procedures. Nevertheless, a simple illustration, for the rcb design mentioned earlier, is instructive. Suppose we have 3 treatments, $t = 3$, and 2 blocks, $b = 2$, and 2 observations per cell, $n = 2$. To simplify things we deal with the no-interaction version of (7):

$$y_{ijk} = \mu + \tau_i + \beta_j + e_{ijk} \,, \tag{26}$$

for $i = 1, 2, 3$, $j = 1, 2$ and $k = 1, 2$. Then $\underset{\sim}{X}$ of (11) is

$$\underset{\sim}{X} = \begin{bmatrix} 1 & 1 & \cdot & \cdot & 1 & \cdot \\ 1 & 1 & \cdot & \cdot & 1 & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot & 1 \\ 1 & 1 & \cdot & \cdot & \cdot & 1 \\ 1 & \cdot & 1 & \cdot & 1 & \cdot \\ 1 & \cdot & 1 & \cdot & 1 & \cdot \\ 1 & \cdot & 1 & \cdot & \cdot & 1 \\ 1 & \cdot & 1 & \cdot & \cdot & 1 \\ 1 & \cdot & \cdot & 1 & 1 & \cdot \\ 1 & \cdot & \cdot & 1 & 1 & \cdot \\ 1 & \cdot & \cdot & 1 & \cdot & 1 \\ 1 & \cdot & \cdot & 1 & \cdot & 1 \end{bmatrix}$$

where a dot represents zero. Going at once to the partitioned form (22) where $\underset{\sim}{X}_1$ is the first 4 columns of $\underset{\sim}{X}$ and $\underset{\sim}{X}_2$ the remaining 3 we find

$$\begin{bmatrix} \underset{\sim}{X}_1'\underset{\sim}{X}_1 & \underset{\sim}{X}_1'\underset{\sim}{X}_2 \\ \underset{\sim}{X}_2'\underset{\sim}{X}_1 & \underset{\sim}{X}_2'\underset{\sim}{X}_2 \end{bmatrix} = \left[\begin{array}{cc:c} 12 & 4\underset{\sim}{1}_3' & 6\underset{\sim}{1}_2' \\ 4\underset{\sim}{1}_3 & 4\underset{\sim}{I}_3 & 2\underset{\sim}{J}_{3x2} \\ \hdashline 6\underset{\sim}{1}_2 & 2\underset{\sim}{J}_{2x3} & 6\underset{\sim}{I}_2 \end{array}\right] \tag{27}$$

where $\underset{\sim}{1}_k$ is a column vector of k unities and $\underset{\sim}{J}_{rxq}$ is an r x q matrix with every element unity. Generalization of (27) gives $\underset{\sim}{X}'\underset{\sim}{X}$ having sub-matrices

$$\underset{\sim}{X}_1'\underset{\sim}{X}_1 = \begin{bmatrix} N & bn\underset{\sim}{i}_a' \\ bn\underset{\sim}{i}_a & bn\underset{\sim}{I}_a \end{bmatrix}, \quad \underset{\sim}{X}_1'\underset{\sim}{X}_2 = \begin{bmatrix} an\underset{\sim}{i}_b' \\ n\underset{\sim}{J}_{axb} \end{bmatrix} \text{ and } \underset{\sim}{X}_2'\underset{\sim}{X}_2 = an\underset{\sim}{I}_b . \tag{28}$$

We now use (24) and (25) to derive $R(\underset{\sim}{\beta}_1|\underset{\sim}{\beta}_2)$ which, as shall be shown, is the treatment sum of squares $SST = bn \sum_{i=1}^{t} (\bar{y}_{i..} - \bar{y}_{...})^2$ of Table 1. First, because we know that in balanced data a solution vector is $\mu^o = \bar{y}_{...}$, $\tau_i^o = \bar{y}_{i..} - \bar{y}_{...}$ and $\beta_j^o = \bar{y}_{.j.} - \bar{y}_{...}$, we know that $\underset{\sim}{\beta}_1^o$ of (22) and (24) is

$$\underset{\sim}{\beta}_1^{o\prime} = [\bar{y}_{...}, \bar{y}_{1..} - \bar{y}_{...}, \cdots, \bar{y}_{t..} - \bar{y}_{...}] . \tag{29}$$

Also, from (28) we get (25) as

$$\underset{\sim}{Q} = \begin{bmatrix} N & bn\underset{\sim}{1}_a' \\ bn\underset{\sim}{1}_a & bn\underset{\sim}{I}_a \end{bmatrix} - \begin{bmatrix} an\underset{\sim}{1}_b' \\ n\underset{\sim}{J}_{a\times b} \end{bmatrix} \frac{1}{an} \underset{\sim}{I}_b \begin{bmatrix} an\underset{\sim}{1}_b & n\underset{\sim}{J}_{b\times a} \end{bmatrix}$$

$$= \begin{bmatrix} N & bn\underset{\sim}{1}_a' \\ bn\underset{\sim}{1}_a & bn\underset{\sim}{I}_a \end{bmatrix} - \begin{bmatrix} abn & bn\underset{\sim}{1}_a' \\ bn\underset{\sim}{1}_a & (bn/a)\underset{\sim}{J}_a \end{bmatrix}$$

$$= \begin{bmatrix} 0 & \underset{\sim}{0} \\ \underset{\sim}{0} & bn\underset{\sim}{I}_a - (bn/a)\underset{\sim}{J}_a \end{bmatrix}.$$

Substituting this and (29) into (24) gives

$$R(\underset{\sim}{\beta}_1 | \underset{\sim}{\beta}_2) = bn \sum_{i=1}^{t} (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2 - (bn/a) \sum_{i=1}^{t} (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot}) \underset{\sim}{1}_a' \begin{bmatrix} \bar{y}_{1\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot} \\ \bar{y}_{t\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot} \end{bmatrix}$$

$$= bn \sum_{i=1}^{t} (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2 ,$$

as anticipated.

This example illustrates how making use of GLM procedures can be a lengthy process. Their starting point is a series of matrix expressions and, in the case of designed experiments, the end point is a set of sums of squares. Intervening algebra is tedious. However, given a linear model and failing an ability to be able to write down an ANOVA table zippity-zap as does the experimental design expert, then the pain of working through the tedious algebra will be worthwhile. It always yields a correct analysis. In contrast, even intelligent guessing on the basis of inexperience, often leads to a wrong analysis.

This is not to suggest that we should not be teaching experimental design; far from it. But what I am suggesting, and emphasizing, is that GLM theory provides back-up procedures that can always be relied on.

Regression For regression we usually have $\underset{\sim}{X}$ of full column rank, so that the partitioned matrix on the left-hand side of (27) has an inverse,

$$\begin{bmatrix} \underset{\sim}{X}_1'\underset{\sim}{X}_1 & \underset{\sim}{X}_1'\underset{\sim}{X}_2 \\ \underset{\sim}{X}_2'\underset{\sim}{X}_1 & \underset{\sim}{X}_2'\underset{\sim}{X}_2 \end{bmatrix}^{-1} = \begin{bmatrix} \underset{\sim}{T}_{11} & \underset{\sim}{T}_{12} \\ \underset{\sim}{T}_{21} & \underset{\sim}{T}_{22} \end{bmatrix}, \text{ say.} \tag{30}$$

Then $\underset{\sim}{Q}$ of (25) is $\underset{\sim}{T}_{11}^{-1}$, so that

$$R(\underset{\sim}{\beta}_1|\underset{\sim}{\beta}_2) = \hat{\underset{\sim}{\beta}}_1\underset{\sim}{T}_{11}^{-1}\hat{\underset{\sim}{\beta}}_1 .$$

This has been referred to as the "invert part of the inverse" rule, special cases of which are available or can be easily derived.

Covariance The partitioned form (21) can be written as

$$\underset{\sim}{y} = \underset{\sim\sim}{X}\underset{\sim}{a} + \underset{\sim\sim}{Z}\underset{\sim}{b} + \underset{\sim}{e} \tag{31}$$

or equivalently as

$$\underset{\sim}{y} = \mu\underset{\sim}{1} + \underset{\sim}{X}_1\underset{\sim}{\alpha} + \underset{\sim\sim}{Z}\underset{\sim}{b} + \underset{\sim}{e} , \tag{32}$$

where

$$\underset{\sim}{X} \equiv [\underset{\sim}{1} \quad \underset{\sim}{X}_1] \quad \text{and} \quad \underset{\sim}{a} \equiv \begin{bmatrix} \mu \\ \underset{\sim}{\alpha} \end{bmatrix} . \tag{33}$$

This dual notation is convenient for considering covariance models: $\underset{\sim}{a}$ is the vector of parameters for factors and interactions including μ, and $\underset{\sim}{\alpha}$ is the corresponding vector without μ. $\underset{\sim}{X}$ is the design matrix associated with $\underset{\sim}{a}$ and $\underset{\sim}{X}_1$ that associated with $\underset{\sim}{\alpha}$. $\underset{\sim}{Z}$ is the matrix of covariable values and $\underset{\sim}{b}$ the

corresponding vector of coefficient parameters or "slopes." We assume, as is reasonable, that $\underset{\sim}{Z}$ has full column rank and its columns are linearly independent of those of $\underset{\sim}{X}$.

It can then be shown that solutions to the normal equations are

$$\underset{\sim}{a}^{*} = (\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{y} - (\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{Z}\hat{\underset{\sim}{b}} \tag{34}$$

and

$$\hat{\underset{\sim}{b}} = (\underset{\sim}{Z}'\underset{\sim}{P}\underset{\sim}{Z})^{-1}\underset{\sim}{Z}'\underset{\sim}{P}\underset{\sim}{y} \tag{35}$$

with

$$\underset{\sim}{P} = \underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}' \ . \tag{36}$$

Also

$$R(\underset{\sim}{\alpha}|\mu) = \underset{\sim}{y}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{y} - N\,\bar{y}_{\cdot}^{2} \ , \tag{37}$$

$$R(\underset{\sim}{b}|\mu) = \underset{\sim}{y}'\underset{\sim}{z}(\underset{\sim}{z}'\underset{\sim}{z})^{-1}\underset{\sim}{z}'\underset{\sim}{y} \ , \tag{38}$$

$$R(\underset{\sim}{b}|\underset{\sim}{a}) = \underset{\sim}{y}'\underset{\sim}{P}\underset{\sim}{Z}(\underset{\sim}{Z}'\underset{\sim}{Z})^{-1}\underset{\sim}{Z}'\underset{\sim}{P}\underset{\sim}{y} \ , \tag{39}$$

and

$$R(\underset{\sim}{\alpha}|\mu,\underset{\sim}{b}) = R(\underset{\sim}{b}|\underset{\sim}{a}) - R(\underset{\sim}{b}|\mu) + R(\underset{\sim}{\alpha}|\mu) \ . \tag{40}$$

Each of these reductions in sums of squares is of some interest. $R(\underset{\sim}{\alpha}|\mu)$ is that for fitting just the factors and interactions part of the model, over and above the mean. $R(\underset{\sim}{b}|\mu)$ is for fitting only the covariates, effectively just a regression model, $\underset{\sim}{z}'\underset{\sim}{z}$ being the matrix of sums of squares and products of the covariates corrected for the mean, and $\underset{\sim}{z}'\underset{\sim}{y}$ the vector of corrected sums of products of the covariates and the y-values. $R(\underset{\sim}{b}|\underset{\sim}{a})$ is for fitting the covariates after accounting for the factors and interactions, a procedure that is often of interest to economists; and conversely, $R(\underset{\sim}{\alpha}|\mu,\underset{\sim}{b})$ is for fitting the factors and interactions after accounting for μ and the covariates—a process

of interest in many biological and other applications.

Reductions $R(\underset{\sim}{\alpha}|\mu)$ and $R(\underset{\sim}{b}|\mu)$ are readily computed since they pertain to familiar factors-and-interactions and to regression models respectively. Properties of $\underset{\sim}{P}$ lead to $R(\underset{\sim}{b}|\underset{\sim}{a})$ being readily computable too, even for unorthodox covariance models with messy data; and then $R(\underset{\sim}{\alpha}|\mu,\underset{\sim}{b})$ is simply a linear combination of the other three, (37), (38) and (39), as shown in (40).

To compute $R(\underset{\sim}{b}|\underset{\sim}{a})$ observe from (36) that for $\underset{\sim}{z}$ being a column of $\underset{\sim}{Z}$, a column of $\underset{\sim\sim}{PZ}$ in (39) is

$$\underset{\sim\sim}{P}\underset{\sim}{z} = \underset{\sim}{z} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'\underset{\sim}{z} = \underset{\sim}{z} - \hat{\underset{\sim}{z}} \qquad (41)$$

where $\hat{\underset{\sim}{z}}$ is the predicted value of $\underset{\sim}{z}$ after fitting, solely for computational purposes, $\underset{\sim}{z} = \underset{\sim\sim}{X}\underset{\sim}{a} + \underset{\sim}{\epsilon}$. Hence $\underset{\sim}{z} - \hat{\underset{\sim}{z}}$ is the vector of estimated residuals which we denote by $\underset{\sim}{r}(\underset{\sim}{z})$ and so have

$$\underset{\sim\sim}{P}\underset{\sim}{z} = \underset{\sim}{r}(\underset{\sim}{z}) \ .$$

It is, of course, invariant to the choice of $(\underset{\sim}{X}'\underset{\sim}{X})^{-}$ used in (41). Carrying this out for each column $\underset{\sim}{z}$ of $\underset{\sim}{Z}$ gives

$$\underset{\sim\sim}{PZ} = \underset{\sim}{R} \ ,$$

a matrix of z-residuals derived from $\underset{\sim}{Z}$ by replacing each of its columns $\underset{\sim}{z}$ by the corresponding vector of residuals $\underset{\sim}{r}(\underset{\sim}{z})$. Then, since $\underset{\sim}{P}$ of (36) is idempotent, (35) and (39) are

$$\hat{\underset{\sim}{b}} = (\underset{\sim}{R}'\underset{\sim}{R})^{-1}\underset{\sim}{R}'\underset{\sim}{y} \quad \text{and} \quad R(\underset{\sim}{b}|\underset{\sim}{a}) = \hat{\underset{\sim}{b}}'\underset{\sim}{R}'\underset{\sim}{y} \ .$$

These expressions are computationally equivalent to fitting the model $\underset{\sim}{y} = \underset{\sim\sim}{R}\underset{\sim}{b} + \underset{\sim}{e}$. Hence for calculating $\underset{\sim}{a}^{*}$, $\hat{\underset{\sim}{b}}$ and $R(\underset{\sim}{b}|\underset{\sim}{a})$ we have the following algorithm:

(1) To each column $\underset{\sim}{z}$ of $\underset{\sim}{Z}$ fit the model

$$\underset{\sim}{z} = \underset{\sim\sim}{X}\underset{\sim}{a} + \underset{\sim}{e} \tag{42}$$

and calculate the vector of residuals

$$\underset{\sim}{r}(\underset{\sim}{z}) = \underset{\sim}{z} - \hat{\underset{\sim}{z}} = [\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}']\underset{\sim}{z} , \tag{43}$$

(2) Replace each column $\underset{\sim}{z}$ of $\underset{\sim}{Z}$ by $\underset{\sim}{r}(\underset{\sim}{z})$ and call the resulting matrix $\underset{\sim}{R}$.

(3) Fit $\underset{\sim}{y} = \underset{\sim\sim}{R}\underset{\sim}{b} + \underset{\sim}{e}$ (a full rank model) and calculate the required terms:

$$\hat{\underset{\sim}{b}} = (\underset{\sim}{R}'\underset{\sim}{R})^{-1}\underset{\sim}{R}'\underset{\sim}{y} ,$$

$$\underset{\sim}{a}^{*} = (\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'(\underset{\sim}{y} - \underset{\sim\sim}{Z}\hat{\underset{\sim}{b}}) ,$$

and

$$R(\underset{\sim}{b}|\underset{\sim}{a}) = \hat{\underset{\sim}{b}}'\underset{\sim}{R}'\underset{\sim}{y} \quad .$$

In this way do we have a mechanism for carrying out all covariance analyses.

Note that the model (42) for $\underset{\sim}{z}$ is exactly the same as the covariance model for $\underset{\sim}{y}$ in (31) but with the covariates omitted. Thus $\underset{\sim}{r}(\underset{\sim}{z})$ of (43) is the same function of $\underset{\sim}{z}$ as $\underset{\sim}{y} - \hat{\underset{\sim}{y}}$ is of $\underset{\sim}{y}$ after fitting the model $\underset{\sim}{y} = \underset{\sim\sim}{X}\underset{\sim}{a} + \underset{\sim}{e}$. This has important ramifications in that for many a covariance model $E(\underset{\sim}{y}) = \underset{\sim\sim}{X}\underset{\sim}{a} + \underset{\sim\sim}{Z}\underset{\sim}{b}$ the factors and interactions part of it, $E(\underset{\sim}{y}) = \underset{\sim\sim}{X}\underset{\sim}{a}$, is well known and so also is the form of the corresponding function of $\underset{\sim}{y}$ that constitutes $\underset{\sim}{r}(\underset{\sim}{y}) = \underset{\sim}{y} - \hat{\underset{\sim}{y}}$. For example, in the covariance rcb model of (8) we know that the non-covariance part of it has $\hat{\underset{\sim}{y}}$ as a vector whose elements are $\bar{y}_{ij\cdot}$; in fact $\hat{\underset{\sim}{y}} = \{\bar{y}_{ij\cdot}\underset{\sim}{1}_n\}$ for $i = 1, \cdots, t$ and $j = 1, \cdots, b$. This means that wherever we know $\underset{\sim}{r}(\underset{\sim}{y})$ for fitting $E(\underset{\sim}{y}) = \underset{\sim\sim}{X}\underset{\sim}{a}$, in order to compute $\underset{\sim}{r}(\underset{\sim}{z})$ in (43) we will not need to calculate $\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-}\underset{\sim}{X}'$ explicitly but from knowing the form of $\underset{\sim}{r}(\underset{\sim}{y}) = \underset{\sim}{y} - \hat{\underset{\sim}{y}}$ can

immediately calculate $\underset{\sim}{r}(\underset{\sim}{z}) = \underset{\sim}{z} - \hat{\underset{\sim}{z}}$. This makes the algorithm's first step very easy.

The covariance example given in (8) is standard and quite straightforward. But consider the following possibility:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + b_1 x_{1,ijk} + b_{2,i} x_{2,ijk} + b_{3,j} x_{3,ijk} + \epsilon_{ijk} .$$

The coefficients of the x_2-covariate differ from treatment to treatment and those of the x_3-covariate differ from block to block. Although there may be difficulty in imagining where such a model equation might be applicable, suppose for the moment that it is; and suppose we have "messy" data, with yields from some treatments missing in some blocks, and with some treatment-block combinations having no data at all. Then even in such an unorthodox case as this the algorithm provides a mechanism for developing the ANOVA tables. And note that $\underset{\sim}{r}(\underset{\sim}{z})$ of the algorithm is easy: $\hat{\underset{\sim}{z}} = \left\{\bar{z}_{ij.}\underset{\sim}{1}_{n_{ij}}\right\}$ for i = 1, ···, t, j = 1, ···, b for $n_{ij} \neq 0$.

μ_{ij}-Models

Difficulties concerning both estimability and restrictions on parameters associated with model equations like (7) have already been mentioned. These difficulties can be quite severe, especially from a teaching and/or interpreting-for-the-layman point of view, particularly when dealing with unbalanced, "messy" or survey data. However, all difficulties vanish if we stay with the model equation $y_{ijk} = \mu_{ij} + \epsilon_{ijk}$ and its more general form given earlier, $y_{ij\cdots pr} = \mu_{ij\cdots p} + \epsilon_{ij\cdots pr}$. Using the 2-factor notation μ_{ij} and $\bar{y}_{ij.}$ generically, and confining attention to only those cells containing data (an eminently "obvious" thing to do), then every μ_{ij} corresponding to a non-zero n_{ij} is estimable

with b.l.u.e. $\hat{\mu}_{ij} = \bar{y}_{ij\cdot}$; and every linear combination of such μ_{ij}'s is estimable and can be the basis for a testable hypothesis. No restrictions on parameters are needed, because there is no overparameterization in the model—there are exactly as many parameters μ_{ij} as there are observed means $\bar{y}_{ij\cdot}$. Thus the whole analysis reduces to being in terms of the observed cell means $\bar{y}_{ij\cdot}$ which are b.l.u.e.'s of the population cell means μ_{ij}.

One consequence of using models of this sort, which we generically call μ_{ij}-models, is that an investigator has to phrase his hypotheses (or the parameters on which he wants confidence intervals) in terms of the population means μ_{ij}. But this is all to the good, because after all, cell means are the crux of one's data.

Cell means were also the effective starting point that Fisher began with. In dealing with designed experiments, as he did, he came up with neat-looking expressions for F-statistics, like those in Table 1. This kind of neatness is not maintained for messy data, but instead we have the universal result that to test the hypothesis $H:\Sigma\Sigma_{ij} p_{ij}\mu_{ij} = m$ we use the F-statistic

$$\frac{\left(\Sigma\Sigma_{ij} p_{ij}\bar{y}_{ij\cdot} - m\right)^2}{\hat{\sigma}^2 \Sigma\Sigma_{ij} p_{ij}/n_{ij}} \sim F_{1,\ N - N'}$$

where there are N observations, N' cells containing data and

$$\hat{\sigma}^2 = \Sigma\Sigma\Sigma_{ijk}\left(y_{ijk} - \bar{y}_{ij\cdot}\right)^2 \Big/ (N - N') \quad .$$

To test a composite hypothesis write the μ_{ij}'s and $\bar{y}_{ij\cdot}$'s in vectors $\underset{\sim}{\mu}$ and $\underset{\sim}{\bar{y}}$ respectively and the corresponding n_{ij}'s in a diagonal matrix $\underset{\sim}{D}\{n_{ij}\}$. Then the hypothesis $H:\underset{\sim}{K}'\underset{\sim}{\mu} = \underset{\sim}{m}$ is tested using

$$(\underset{\sim}{K}'\bar{\underset{\sim}{y}} - \underset{\sim}{m})'(\underset{\sim}{K}'\underset{\sim}{D}\{1/n_{ij}\}\underset{\sim}{K})^{-1}(\underset{\sim}{K}'\bar{\underset{\sim}{y}} - \underset{\sim}{m})/s\hat{\sigma}^2 \sim F_{s,N-N'} ,$$

where $\underset{\sim}{K}'$ has full row rank $s < N'$.

There are encouraging signs in the literature that these simple and powerful μ_{ij}-models are starting to gain favor, e.g. Hocking and Speed [1972], Urquhart *et al.* [1973] and Kutner [1974]. Hopefully their wider adoption will continue.

Conclusions

General linear model theory provides procedures for deriving ANOVA methodology not only for complicated experimental designs but also for survey data, for regression, and for covariance with any number of covariates in either experimental or survey data. GLM procedures therefore deserve to be emphasized. and utilized much earlier in our teaching programs than is presently the custom. Without prior teaching of matrix algebra it would certainly be difficult to teach GLM procedures, which possibly precludes their being taught at the level of a general methods course. However, serious thought could well be given to developing something at this level. The unifying features and wide embrace of GLM procedures show the student that many things he may otherwise learn as being unrelated are nothing more than special cases of just one aspect of statistics. And emphasis can also be made that analysis of variance is indeed just one aspect of statistics.

Students and practitioners also derive from GLM procedures the facility to develop ANOVA methods for any linear model, safe in the knowledge that, for any postulated linear model, the correct analysis will be obtainable. And finally, by increased adoption of the μ_{ij} models we will dispel all the difficulties currently associated with overparameterization and estimability.

References

Hocking, R. R. and F. M. Speed [1972]. A full rank analysis of some linear model problems. [Submitted to J. Amer. Statist. Assoc.]

Kutner, M. H. [1974]. Hypothesis testing in linear models (Eisenhart Model 1). Amer. Statistician 28, 98.

Lum, Mary D. [1954]. Rules for determining error terms in hierarchal and partially hierarchal models. Wright Air Development Center, Dayton, Ohio.

Schultz, E. F., Jr. [1955]. Rules of thumb for determining expectations of mean squares in analysis of variance. Biometrics 11, 123-135.

Searle, S. R. [1971]. Linear Models, Wiley, N. Y.

Urquhart, N. S., D. L. Weeks and C. R. Henderson [1972]. Estimation associated with linear models: a revisitation. Communications in Statistics 1, 303-330.